DISCLAIMER

Computation of Small Angle Scattering profiles with 3D Zernike Polynomials

Haiguang Liu,[a] Richard J. Morris,[b] Alexander Hexemer,[c]
Scott Grandison[d] and Peter H. Zwart [a]*

[a]*Physical Bioscience Division, Lawrence Berkeley National Laboratories, Berkeley, CA, USA,* [b]*Computational System Biology, John Innes Centre, Norwich Research Park, Norwich NR47UH, UK,* [c]*Advanced Light Source, Lawrence Berkeley National Laboratories, Berkeley, CA, USA, and* [d]*School of Computing Sciences, University of East Anglia, Norwich NR47TJ, UK. E-mail: PHZwart@lbl.gov*

(Received 0 XXXXXXX 0000; accepted 0 XXXXXXX 0000)

Abstract

Small angle scattering methods are extensively used for characterizing macromolecular structure and dynamics in solution. The computation of theoretical scattering profiles from 3D models is crucial in order to test structural hypotheses. Here, we present a new approach to efficiently compute SAXS profiles that are based on 3D Zernike polynomial expansions. Comparison to existing methods and experimental data shows that the Zernike method can be used to effectively validate 3D models against experimental data. For molecules with large cavities or complicated surfaces, the Zernike method more accurately accounts the solvent contributions. The program is available as open source software at *http://sastbx.als.lbl.gov*.

1. Introduction

Knowledge of the three dimensional structures of macromolecules provides essential insights into biology at an atomic level (Orengo *et al.*, 1999). The vast majority of macromolecular structures are determined by X-ray crystallography, typically providing high-resolution models with sub-Ångstrom precision in the refined atomic coordinates. Unfortunately, not all proteins crystalize and, more often than not, the structure and dynamics of proteins in solution are quite different than what is observed in the crystal (Glatter & Kratky, 1982). The behaviour of macromolecules in solution can studied using Small Angle X-ray Scattering (SAXS) (Hura *et al.*, 2009; Koch *et al.*, 2003; Stuhrmann, 2008). Although SAXS is a low resolution technique, typically only providing scattering data from 30 to 20 Ångstrom, the data can be interpreted with the aid of known crystal structures. The synergistic use of high-resolution atomic models in combination with SAXS data can result in a fundamental comprehension of the biological relevance of molecules in a near-native environment (Grishaev *et al.*, 2005; Putnam *et al.*, 2007; Wang *et al.*, 2008; Grant *et al.*, 2011).

1.1. Orientational averaging

The calculation of SAXS profiles can be carried out using the Debye formula (Debye, 1915) with explicit (Grishaev *et al.*, 2010; Durchschlag & Zipper, 2003) or implicit (Schneidman-Duhovny *et al.*, 2010; Poitevin *et al.*, 2011) modelling of border-bound and excluded solvent. The main problem with the Debye method is the computational complexity: for each value of momentum transfer q ($q = 4\pi sin(\theta)/\lambda$, where λ is the wavelength and 2θ the scattering angle), a double summation of order N^2 needs to be carried out, where N is equal to the number of atoms. Distance binning procedures and other techniques (Stovgaard *et al.*, 2010; Schneidman-Duhovny *et al.*, 2010; Tjioe & Heller, 2007) can reduce the complexity significantly but still suffer from difficulties

associated with modeling bound and excluded solvent.

The best known numerical procedure to reduce the computational complexity from $O(N^2)$ to $O(N)$ is the spherical harmonics expansion (SHE) originally proposed by Stuhrmann and Svergun and implemented in the program CRYSOL (Stuhrmann, 1970*b*; Svergun *et al.*, 1995).

The SAXS intensity can be calculated as:

$$I(q) = \langle |A_{atoms}(\mathbf{q}) - \rho_0 A_{excl}(\mathbf{q}) + \delta\rho_0 A_{bound}(\mathbf{q})|^2 \rangle_\Omega \tag{1}$$

where the complex quantity A is the Fourier transform the electron density from the particle (subscript *atoms*), excluded solvent (subscript *excl*) and surface bound solvent (*bound*) respectively. The averaging in the above expression is carried out over the solid angle Ω and $\mathbf{q} = (q, \Omega)$.

The averaging over the solid angle can be carried out in several ways. First of all, one can choose the route adopted by Debye in which the orientational average of the complex exponent is evaluated analytically to be a *sinc* function. For clarity, the expressions below only contain atomic contributions:

$$A_{atoms}(\mathbf{q}) = \sum_{j=1}^{N} f_j(q) \exp\left[-i\mathbf{q}\mathbf{r_j}\right] \tag{2}$$

and thus

$$I_{atoms}(q) = \sum_{j=1}^{N}\sum_{k=1}^{N} f_j(q) f_k(q) \int_\Omega \exp\left[-i\mathbf{q}(\mathbf{r_j} - \mathbf{r_k})\right] d\Omega \tag{3}$$

the latter integral evaluates as a sinc function such that one obtains

$$I_{atoms}(q) = \sum_{j=1}^{N}\sum_{k=1}^{N} f_j(q) f_k(q) \frac{sin(qr_{jk})}{qr_{jk}} \tag{4}$$

The complexity of the above expression is $O(N^2)$.

Instead of evaluating the complex exponent analytically, one can approximate it with a series expansion containing Bessel functions and spherical harmonics (Edmonds,

1957):

$$A_{atoms}(\mathbf{q}) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{+l} 4\pi i^l Y_{lm}(\Omega) \sum_{j}^{N} f_j(q) j_l(qr_j) Y^*_{lm}(\omega_j) \tag{5}$$

where Y_{lm} is a spherical harmonic of order *(l,m)*, (r_j, ω_j) the polar coordinates of atom j and $f_j(q)$ the atomic scattering factor and j_l a spherical bessel function of order l. Setting

$$a_{lm}(q) = \sum_{j=1}^{N} f_j(q) j_l(qr_j) Y^*_{lm}(\omega_j) \tag{6}$$

one obtains

$$A_{atoms}(\mathbf{q}) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{+l} 4\pi i^l a_{lm}(q) Y_{lm}(\Omega) \tag{7}$$

Subsequent averaging over the solid angle is now greatly simplified by the orthogonality properties of spherical harmonics (Edmonds, 1957), resulting in

$$I_{atoms}(q) = 16\pi^2 \sum_{l=0}^{l_{max}} \sum_{m=-l}^{+l} |a_{lm}(q)|^2 \tag{8}$$

As is clear from the above expression, the complexity is reduced from $O(N^2)$ to $O(N)$, because the costly double summation used in the Debye equation is replaced by a single summation for each index *(l,m)*.

1.2. Excluded and surface-bound solvent

The surface-bound solvent can be modeled in various ways. Firstly, the method proposed by Stuhrmann (Stuhrmann, 1970*a*) introduces a single uniform solvent layer around the macromolecule using a two dimensional angular function $F(\omega)$. The advantage of this method is its numerical simplicity in generating the scattering amplitudes from this border layer, involving precomputed partial integrals of spherical bessel functions. Another approach is to use the modified scattering factor approach which includes modeling of excluded solvent and possible surface bound solvent (Schneidman-Duhovny *et al.*, 2010):

$$f_j(q) = f_v(q) - c_1 f_s(q) + c_2 s_i f_w(q) \tag{9}$$

where $f_v(q)$ is equal to the atomic form factor *in vacuo*, $f_s(q)$ the form factor of a dummy atom representing the excluded solvent, s_i the solvent accesibility of the atom and f_w the form factor of water. Coefficients c_1 and c_2 model density of excluded solvent and bound surface water respectively.

The drawback of the dummy atom approach for modeling displaced solvent is that non-uniformities in the density can be introduced by overlapping dummy atoms or empty spaces where in reality one would expect a continuum of uniform solvent. These non-uniformities typically do not have significant effects on the scattered intensities for small values of momentum transfer. The Stuhrmann approach of introducing a uniform layer around the macromolecule can also be problematic. For proteins containing cavities or those with a non-star shape, the uniform layer around the convex hull of the protein will introduces artificial areas without any density. For proteins like chaperonins, the inner surfaces could not be modelled with the Struhrmann approach.

An alternative route for taking into account excluded and surface-bound solvent is by explicit real space modeling of these moieties (Grishaev *et al.*, 2010). A thorough approach is to add the solvation layer using molecular modeling techniques (Park *et al.*, 2009). The main drawback of this route is the computational effort involved in building the explicit solvent model. Another approach is found in the so-called (modified) cube-method (Bardhan *et al.*, 2009). The cube method for modeling excluded solvent is reminiscent of modeling bulk solvent in macromolecular crystallography where the Fourier Transform of a binary mask modeling for the excluded and surface bound solvent was used (Jiang & Brunger, 1994).

In this communication, an approach related to the cube method is developed. In the **Method** section, the detailed derivation and the procedure parametrizing three-dimensional bodies via a 3D Zernike expansion are summarized. Following that, in the **Results** section, the computed SAXS profiles are compared to the results obtained

using the spherical harmonics expansion method. The fitting to a set of experimental data shows that the method can be used to validate 3D models against SAXS experimental data. The advantages of the Zernike method are discussed.

2. Methods

2.1. Zernike Polynomials

3D Zernike polynomials are natural extensions of 2D Zernike polynomials into the third dimension. Basic properties and theory are reviewed in detail elsewhere (Canterakis, 1999). Here, a brief summary is provided. A 3D Zernike polynomial $Z_{nlm}(\mathbf{r})$ is defined as

$$Z_{nlm}(\mathbf{r}) = R_{nl}(r)Y_{lm}(\omega) \tag{10}$$

Where

$$R_{nl}(r) = \sum_{k=0}^{(n-l)/2} N_{nlk} r^{n-2k} \tag{11}$$

$$N_{nlk} = (-1)^k 2^{l-n} \sqrt{2n+3} \times \frac{(2n-2k+1)![(1/2)(n+l)-k]!}{[(1/2)(n-l)-k]!(n+l-2k+1)!(n-k)!k!} \tag{12}$$

and $Y_{lm}(\omega)$ is a spherical harmonic. The order indices must satisfy the following conditions: $n \geq l$, and (n-l) is even; $-l \leq m \leq l$. Zernike polynomials are orthogonal functions on the unit ball:

$$\int_{r\leq 1} Z_{nlm}(\mathbf{r}) Z_{n'l'm'}(\mathbf{r}) d\mathbf{r} = \delta_{nn'}\delta_{ll'}\delta_{mm'} \tag{13}$$

In light of the above orthogonality properties, any twice-differentiable function on the unit ball can be expanded in a series of 3D Zernike polynomials:

$$\rho(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{l=0}^{n} \sum_{m=-l}^{+l} c_{nlm} Z_{nlm}(\mathbf{r}) \tag{14}$$

The complex expansion coefficients c_{nlm}, also known as 3D Zernike moments, can be obtained using the Novotni & Klein algorithm (Novotni & Klein, 2003).

The Fourier transform of $\rho(\mathbf{r})$ parametrized by a Zernike expansion can be derived in a straightforward manner:

$$\begin{aligned}\mathcal{F}\left[\rho(\mathbf{r})\right] &= \int_{r\leq 1} \rho(\mathbf{r}) \exp\left[i\mathbf{qr}\right] d\mathbf{r} &(15)\\ &= \sum_{n=0}^{\infty}\sum_{l=0}^{n}\sum_{m=-l}^{+l} c_{nlm}\mathcal{F}\left[Z_{nlm}(\mathbf{r})\right] &(16)\end{aligned}$$

$$\exp\left[i\mathbf{qr}\right] = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l} i^l j_l(qr) Y_{lm}(\omega_q) Y^*_{lm}(\omega_r) \quad (17)$$

Fourier transform of a single Zernike polynomial is then (Mathar, 2008)

$$\begin{aligned}\mathcal{F}\left[Z_{nlm}(\mathbf{r})\right] &= 4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} i^{l'}\int_0^1 j_{l'}(qr) R_{nl}(r) r^2 dr \times \\ &\quad Y_{l'm'}(\omega_q)\int_{\omega_r} Y^*_{l'm'}(\omega_r) Y_{lm}(\omega_r) d\omega_r \\ &= 4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} i^{l'} \times \\ &\quad \int_0^1 j_{l'}(qr) R_{nl}(r) r^2 dr Y_{l'm'}(\omega_q)\delta_{ll'}\delta_{mm'} \\ &= 4\pi i^l Y_{lm}(\omega_q)\int_0^1 j_l(qr) R_{nl}(r) r^2 dr \\ &= 4\pi i^l Y_{lm}(\omega_q)\frac{j_n(q)+j_{n+2}(q)}{2n+3}(-1)^{(n-l)/2} \\ &= 4\pi i^l (-1)^{(n-l)/2} Y_{lm}(\omega_q) b_n(q) &(18)\end{aligned}$$

and thus

$$A(\mathbf{q}) = 4\pi\sum_{n=0}^{\infty}\sum_{l=0}^{n}\sum_{m=-l}^{+l} i^l (-1)^{(n-l)/2} c_{nlm} Y^*_{lm}(\omega_q) b_n(q) \quad (19)$$

with

$$b_n(q) = \frac{j_n(q)+j_{n+2}(q)}{2n+3} \quad (20)$$

2.2. SAXS intensity

SAXS curves are equal to the spherically averaged squared moduli of the Fourier transform of the scattering object:

$$I(q) = \int_{\Omega} A(\mathbf{q}) A^*(\mathbf{q}) d\omega_q \tag{21}$$

Given the orthogonality properties of spherical harmonics the above expression reduces to

$$\begin{aligned} I(q) &= 16\pi^2 \sum_{n=0}^{\infty} \sum_{n'=0}^{\infty} b_n(q) b_{n'}(q) \times \\ &\quad \sum_{l=0}^{n} (-1)^{(n+n')/2-l} \sum_{m=-l}^{+l} c_{nlm} c^*_{n'lm} \end{aligned} \tag{22}$$

Note that all expressions above still assume a particle of unity radius. With r_{max} the particle radius, the above expression can be modified to include the particle size. With an additional regrouping of constants, an economical expression for the SAXS intensity is obtained:

$$I(q) = 16\pi^2 \sum_{n=0}^{\infty} \sum_{n'=0}^{\infty} b_n(qr_{max}) b_{n'}(qr_{max}) F_{nn'} \tag{23}$$

$$\begin{aligned} F_{nn'} &= \sum_{l=0}^{n} k_{nn'l} \sum_{m=-l}^{+l} c_{nlm} c^*_{n'lm} \\ k_{nn'l} &= (-1)^{(n+n')/2-l}. \end{aligned}$$

2.3. Real Space Modeling

As described by Novotni & Klein and Mak *et al*, Zernike moments from 3D bodies can be efficiently obtained via a linear combination of geometric moments of the object (Novotni & Klein, 2003; Mak *et al.*, 2008):

$$c_{nlm} = \frac{3}{4\pi} \sum_{r+s+t \leq n} \overline{\chi^{rst}_{nlm}} M_{rst} \tag{24}$$

where M_{rst} is the geometric moment

$$M_{rst} = \int_{|\mathbf{r}|\leq 1} \rho(\mathbf{r}) x^r y^s z^t d\mathbf{r} \tag{25}$$

which can be computed from the voxelized object. The detailed procedure to compute coefficients χ_{nlm}^{rst} has been outlined by Nototni & Klein (Novotni & Klein, 2003). The voxelization procedure maps a continuous electron density onto a discrete collection of voxels from which the Zernike moments are computed. The voxelized electron density forms the basis for the bound and excluded solvent model. Representing the set of non-zero electron density voxels with $\mathbf{P}$, the set of voxels representing the excluded solvent and the surface bound solvent ($\mathbf{S} + \mathbf{B}$) is obtained by masking the voxels within 3.0 Å of atoms. The set of voxels of the excluded solvent $\mathbf{S}$ is obtained by removing elements in the set $\mathbf{S} + \mathbf{B}$ that lie within 3.0 Å of the surface (via erosion procedure shown in Figure 1). The benefit of this operation over a dummy atom approach is that one avoids introduction of overlaps and non-physical voids in the excluded solvent. Furthermore, the approach outlined will provide a border layer for all solvent exposed surfaces, including large voids. A graphical representation of the above procedure is shown in Figure 1. In order to compute the final scattering curves, the zernike moments of three voxelized objects are weighted by appropriate contrast levels, summed and result in a single set of zernike moments from which a scattering curve can be obtained via equation (23). It is worthwhile to mention that the three voxelized objects, $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{B}$, are scaled down by the same r_{max} to fit in the unit sphere for the calculation of the corresponding moments. The same r_{max} is used for the SAXS profile calculation, as shown in Eqn. (23).

2.4. SAXS profile fitting and comparison

The discrepancy between the experimental data and the computed theoretical SAXS profile is measured using the χ^2 scoring function, defined as:

$$\chi^2 = \frac{1}{N_{obs}} \sum_{j=1}^{N_{obs}} \left[\frac{I_{obs}(q_j) - kI_{calc}(q_j) + c}{\sigma_j} \right]^2 \tag{26}$$

where the factor k is a scaling factor and c is the background correction, both of which can be obtained from standard least square fitting (Lawson & Hanson, 1987). In this work, the χ is reported when fitting to the experimental data.

3. Results

3.1. Basic identities

Two basic SAXS invariants can be readily derived from expression 23. First of all, the total forward scattering, I(0) is equal to $|c_{000}|^2$ since $j_n(0) = 0$ for $n > 0$. Furthermore, expanding I(q) around 0 by truncating expression 23 to an order of 0 and assuming a mean density of unity, one obtains

$$I(q) \propto \frac{(j_0(qr_{max}) + j_2(qr_{max}))^2}{9} \tag{27}$$

It can be easily seen that the above expression is equivalent to the scattering of a solid sphere (Glatter & Kratky, 1982).

The equation 23 can be expanded around 0 to get the following expression:

$$I(q) \propto 1 - \frac{q^2 r_{max}^2}{5} + O(q^4 r_{max}^4) \tag{28}$$

In the very small q-region, the higher order terms are neglectable and this expression thus reduces to the guinier approximation. Using radius of gyration (R_g) for a solid

sphere, one obtain the familiar Guinier approximation:

$$R_g = \sqrt{3/5}\, r_{max} \tag{29}$$

$$I(q) \propto 1 - \frac{q^2 {R_g}^2}{3} \tag{30}$$

$$\propto \exp[-\frac{q^2 {R_g}^2}{3}] \tag{31}$$

3.2. In vacuo particle scattering

To validate the described derivation, SAXS intensity profiles for particles *in vacuo* are computed via the Debye equation (Debye, 1915) as well as using expression 23. As shown in Figure 2, for lysozyme protein at the maximum expansion order of 20, the SAXS intensity up to a momentum transfer of 0.5 $\mathring{A}^{-1}$ is in perfect agreement with the intensity as computed via the Debye equation. At a momentum transfer value of larger than 0.5 $\mathring{A}^{-1}$ truncation ripples are observed which are dominated by the spherical Bessel function of order 0. When we increase the maximum expansion order to 40, the calculated intensity profiles using both methods agree with each other up to 1.0 $\mathring{A}^{-1}$. For typical sized globular proteins (100∼400 residues), the default maximum expansion order of 30 (i.e., n_{max}=30) works well for small angle scattering region (q <0.5 $\mathring{A}^{-1}$). For elongated models, typically higher expansion orders are recommended as compared to more globular shapes at any q-value. As shown in Figure 2, the higher order polynomials contribute to larger momentum transfer regions only, so a good choice of n_{max} is a balance of accuracy and computational speed. Because the majority of SAXS profiles available do not exceed $q = 0.4\mathring{A}^{-1}$, a default expansion order of 30 is sufficient for most applications.

3.3. Hydrated particle scattering

Since small angle X-ray scattering experiments for biomolecules are conducted in solution to study their structures and dynamics, the scattering *in vacuo* is not appro-

priate to be compared to the solution SAXS data. The contributions of solvents originate from two sources: (a) the excluded solvent; and (b) the (partially) ordered solvent at the surfaces of the molecules resulted denser electron density (compared to the electron density in bulk solvent). This has been elaborated in detail by Svergun and coworkers (Svergun *et al.*, 1995). We computed a few representative macromolecule SAXS profiles using an in house implemented spherical harmonics expansion based method (SHE) and the presented Zernike method (ZNK). The SAXS profiles from the individual components are compared correspondingly, as summarized in Figure 3.

As discussed in the previous section, the computed theoretical scattering profiles using both methods are in excellent agreements for the protein *in vacuo* (**P**). The results shown in Figure 3 also indicate that the scattering profiles for the excluded solvent (**S**) are very similar despite different approaches of solvent modeling. The striking difference is observed for the scattering contribution from the surface bound solvent (**B**), shown as lines (SHE) and circles (ZNK) in blue color in Figure 3. Figure 3A shows lysozyme (PDB ID: 6LYZ) and the scattering profiles, where two methods gets similar profiles for the bound solvent. For proteins with more complicated surfaces, reflected by more bumpy surfaces, the differences in scattering profiles calculated using SHE and ZNK can be found. A typical example is shown in Figure 3B, a myosin domain (PDB ID: 3PN7), for which SHE and ZNK result different scattering profiles for the bound solvent. Furthermore, the spherical harmonics expansion is limited by only modeling the outer surface layer. For molecules with large cavities or holes the inner surface is completely neglected using this approach. The resulting scattering profiles for the border layer are significantly different for such proteins, as shown in Figure 3C (PDB ID: 2E2G).

3.4. Fitting to experimental data

The calculated SAXS profiles can be fit to experimental data to validate the 3D models. To improve the fit to experimental data, the Zernike method can optimize the bound solvent contrast level ($\delta\rho_0$). The Zernike method adopts the approach taken by the Svergun and coworkers (Svergun *et al.*, 1995) to obtain the optimal contrast level: the program scans contrast levels within a predefined range ([0.0, 0.09] e/$\mathring{A}^3$). The average electron density of the excluded solvent is set to be 0.334 e/$\mathring{A}^3$ for typical SAXS experiments, which can be changed to match particular experimental setups.

To test the describe procedure, ten high-quality SAXS data sets from Grant *et al.* (Grant *et al.*, 2011) were used. The data sets used had high resolution crystal structures available compromising over 90% of the particle used in the SAXS studies. Figure 4 summarizes the plots of the SAXS profiles, where the theoretical curves are calculated from the high resolution crystal structures and fitted to the corresponding experimental data. In all the calculations, we used $n_{max} = 30$ for ZNK method and $l_{max} = 15$ for SHE method. The corresponding χ scores are summarized in Table 2. Both the results shown in Figure 4 and Table 2 demonstrate computed SAXS profiles agree to experimental data.

4. Discussion

The Debye formula and variants have advantages of easy implementation, but they do not scale well with system size. For a molecule with **N** atoms and **M** desired intensity data points, the computational complexity is $O(MN^2)$. The spherical harmonics expansion approach improves the complexity to linear w.r.t. number of atoms **N**, giving $O(MN)$. However, the computation time still linearly depends on the number of data points, **M**. In the Zernike method, the position of atoms (x, y, z) and the momentum transfer q are decoupled (see, equations 18-23), and thus the com-

putational complexity is reduced to O(N) with some overhead to generate voxelized objects. Therefore, the presented Zernike method has speed advantages when large numbers of data points are desired (Figure 5).

Zernike expansion method can model the holes/cavities of macromolecules that are usually not well handled in the spherical harmonics expansion methods. Usually, more polynomials are required to meet the desire of resolution and accuracy (see Figure 2). If the desired q-range is up to 0.5 $\mathring{A}^{-1}$, the maximum expansion order n_{max} should be not less than 30 for typical macromolecules. It is worthwhile to note that the number of Zernike polynomials escalates cubicly w.r.t. the maximum expansion order n_{max} (see Table 1), whereas the number of $F_{nn'}$ coefficients increases quadratically. When the high q data is not desired or not available, the execution time can be significantly reduced by using smaller n_{max}.

Even though the described method for construction surface bound solvent is more appropriate than the single outer surface method, the uniform body approach used is not sufficient to describe features seen at high resolution (Bardhan *et al.*, 2009; Park *et al.*, 2009), due to the average internal structure of the solvent shell. The approach described here aims to provide model data at modest scattering angles ($< 0.5\mathring{A}^{-1}$) allowing us to ignore internal structure in the surface bound solvent model. In the WAXS regime, the explicit solvent molecules are necessary to model the scattering profile more accurately (Park *et al.*, 2009).

The spherical harmonics expansion method has been used widely to compute SAXS profiles for comparison with experimental data. Here, the Zernike expansion method provides an alternative way of modeling the excluded solvent and the molecular surface-bound solvent. As described in the **Method** section, this approach models complicated surface bound solvents more accurately. It has been pointed out that the dummy atom approach using in the spherical harmonics expansion method causes

inaccurate modeling of the excluded solvent by introducing overlaps and gaps between the dummy atoms. A more appropriate way of estimating excluded solvent is to use the union of the dummy atoms, however, the gaps are still to be filled (Bardhan *et al.*, 2009). The treatment of excluded solvent as a uniform density body better reflects the small angle scattering characteristics of bulk solvent, which is due to the fact that water molecules are randomly oriented and spherically averaged to get the SAXS profile. A properly selected splat-range (see Figure1) ensures that no gaps are in the molecular interior, while using the uniform density for the 'marked' voxels guarantees the excluded solvent does not overlap that is often observed in the dummy atom approach. This voxelization approach follows the idea proposed by Bardhan *et al.* that avoids gaps and overlaps introduced in dummy solvent atom approach (Bardhan *et al.*, 2009).

The surface-bound solvent scattering profile differences are due to different ways of modeling molecular surfaces. In the SHE approach, the surface is represented by a set of vectors pointing outwards from the center, whose directions are picked to uniformly sample the points on the surface of a sphere. For small globular molecules, it is reasonable to assume the smooth continued surfaces and the single uniform border layer yields results that agree with the Zernike method. For molecules with more complicated surfaces, a larger number of surface points are required to model possible large curvatures using the uniform outer layer approach. For proteins exhibiting cavities, the single outer surface layer is insufficient for modeling surface bound solvent. As described in the **Method** section, the Zernike method uses voxelized representations of the scattering density, effectively circumventing problems associated with the single bound solvent layer approach, such that the solvent layers surrounding the cavities or holes in the molecules are effectively modeled. The proposed Zernike expansion approach takes all solvent accessible surfaces into account, therefore, the contribu-

tions of the surface bound solvents are correctly incorporated to the overall scattering profiles.

When comparing calculated profiles using the SHE and ZNK methods to the test data, we can see that the both procedures give very similar results. The spherical harmonics expansion method achieves better fits in terms of smaller χ scores for some of the data. It is probably due to the fact that the spherical harmonics expansion methods has more refine-able parameters such as the average excluded volume per atomic group. In the Zernike model, the excluded solvent is treated as a uniform continuous object, therefore there is no further optimization for the excluded solvent scattering at the present time. The bound solvent contrast layer is more relevant to the overall scattering intensity in solution. By optimizing only the latter parameter, one can keep model simpler and significantly reduce the risk of over-fitting.

5. Conclusion

Modelling excluded and bound solvent of macromolecular models in order to calculate accurate theoretical SAXS profiles presents a computational challenge. A new method based on a 3D Zernike polynomial expansion is presented. This method treats excluded solvent as a continuous, uniform density, object and is capable of modeling complicated bound-solvent layers. The results agree for simple shapes with theoretical results calculated using the spherical harmonics expansion method. For molecules with complicated surfaces, the Zernike method offers a natural extension that may help improve the fitting to experimental data. The program and source code, as well as an online webserver, are freely available from *http://sastbx.als.lbl.gov*.

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Bardhan, J., Park, S. & Makowski, L. (2009). *J Appl Crystallogr*, **42**(5), 932–943.

Canterakis, N. (1999). In *Scandinavian Conference on Image Analysis*.

Debye, P. (1915). *Annalen der Physik*, **351**(6), 809–823.

Durchschlag, H. & Zipper, P. (2003). *Eur Biophys J*, **32**(5), 487–502.

Edmonds, A. R. (1957). *Angular Momentum in Quantum Mechanics*. Princeton Univ. Press.

Glatter, O. & Kratky, O. (eds.) (1982). *Small Angle X-ray Scattering*. London: Academic Press.

Grant, T. D., Luft, J. R., Wolfley, J. R., Tsuruta, H., Martel, A., Montelione, G. T. & Snell, E. H. (2011). *Biopolymers*, **95**(8), 517–530.

Grishaev, A., Guo, L., Irving, T. & Bax, A. (2010). *Journal of the American Chemical Society*, **132**(44), 15484–15486.

Grishaev, A., Wu, J., Trewhella, J. & Bax, A. (2005). *J Am Chem Soc*, **127**(47), 16621–8.

Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, 2nd, F. L., Tsutakawa, S. E., Jenney, Jr, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W. & Tainer, J. A. (2009). *Nat Methods*, **6**(8), 606–12.

Jiang, J. & Brunger, A. (1994). *Journal of Molecular Biology*, **243**(1), 100–115.

Koch, M. H., Vachette, P. & Svergun, D. I. (2003). *Q Rev Biophys*, **36**(2), 147–227.

Lawson, C. & Hanson, R. (1987). *Solving Least-Square Problems*. Philadelphia: Society for Industrial and Applied Mathematics.

Mak, L., Grandison, S. & Morris, R. J. (2008). *J Mol Graph Model*, **26**(7), 1035–45.

Mathar, R. (2008). *Baltic Astronomy*, **17**(3/4), 383–398.

Novotni, M. & Klein, R. (2003). In *Proceedings of the eighth ACM symposium on Solid modeling and applications*, SM '03, pp. 216–225. New York, NY, USA: ACM.

Orengo, C. A., Todd, A. E. & Thornton, J. M. (1999). *Current Opinion in Structural Biology*, **9**(3), 374 – 382.

Park, S., Bardhan, J. P., Roux, B. & Makowski, L. (2009). *J Chem Phys*, **130**(13), 134114.

Poitevin, F., Orland, H., Doniach, S., Koehl, P. & Delarue, M. (2011). *Nucleic Acids Res*. Published online.

Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Quarterly Reviews of Biophysics*, **40**(03), 191–285.

Schneidman-Duhovny, D., Hammel, M. & Sali, A. (2010). *Nucleic Acids Res*, **38**, W540–4.

Stovgaard, K., Andreetta, C., Ferkinghoff-Borg, J. & Hamelryck, T. (2010). *BMC Bioinformatics*, **11**, 429.

Stuhrmann, H. B. (1970*a*). *Z. Phys. Chem. Frankfurt*, **72**, 177–184, 185–198.

Stuhrmann, H. B. (1970*b*). *Acta Crystallographica Section A*, **26**(3), 297–306.

Stuhrmann, H. B. (2008). *Acta Crystallographica Section A*, **64**(1), 181–191.

Svergun, D., Barberato, C. & Koch, M. (1995). *Journal of Applied Crystallography*, **28**, 768–773.

Tjioe, E. & Heller, W. T. (2007). *Journal of Applied Crystallography*, **40**, 782–785.

Wang, Y., Trewhella, J. & Goldenberg, D. P. (2008). *J Mol Biol*, **377**(5), 1576–92.

Table 1. *n_{max} and the corresponding number of parameters in the Zernike based method. The number of expansion coefficients c_{nlm} is approximately proportional to $n_{max}^{2.5}$. The number of coefficients needed to compute a SAXS curve is approximately proportional to $(n_{max}/2+1)^2$*

n_{max}	number of c_{nlm}	number of $F_{nn'}$
10	286	36
11	364	42
12	455	49
13	560	56
14	680	64
15	816	72
16	969	81
17	1140	90
18	1330	100
19	1540	110
20	1771	121
21	2024	132
22	2300	144
23	2600	156
24	2925	169
25	3276	182
26	3654	196
27	4060	210
28	4495	225
29	4960	240
30	5456	256

Table 2. *Fit to experimental data comparing the results from Spherical Harmonic expansion (SHE) and the Zernike based method (ZNK) to selected data set from Grant et al. (Grant* et al.*, 2011)*

DATA ID	SHE FIT	ZNK FIT
2	3.0	2.9
3	1.6	1.6
5	2.4	2.7
8	2.1	2.0
9	3.5	3.0
10	3.8	3.4
13	1.4	1.6
14	2.3	2.4
15	1.9	2.1
16	2.0	2.3

Fig. 1. The voxelized representations of protein and solvents. **A**. A molecule will be mapped to 3D voxelized objects to model protein (**P**) and Solvent (**S**+**B**), and the surface bound solvent (**B**) can be decoupled from the excluded solvent (**B**) by erosion operations. **B**. The splat_range and the erosion range are defined as in the 2D scheme plot.

Fig. 2. The SAXS profile for lysozyme *in vacua*. **A**. The resolution of the models in q-space is determined by the maximum expansion order (n_{max}): using n_{max}=40, the intensity can be accurately computed to 1.0 $\mathring{A}^{-1}$. **B**. The corresponding reconstructed models are shown at n_{max} = 10,20,30 and 40.

Fig. 3. The SAXS profiles of individual components. The SHE (solid lines) and ZNK (open circles) method give very similar results for the overall scattering profiles (colored in black). The scattering profiles for the protein (*in vacua*) (red) and excluded solvent (green) from two methods are also in good agreements. For molecules with irregular surfaces (**B**) or large cavities/holes (**C**), the surface bound solvents have very different scattering profiles (blue). The default parameters are used: l_{max}=15 for SHE and n_{max}=30. The corresponding PDB IDs are 6LYZ(**A**), 3PN7(**B**), and 2E2G(**C**).

Fig. 4. The fit to experimental data. The red curves are calculated using SHE method and the green curves are computed using zernike method. The default parameters are used: l_{max}=15 for SHE, and n_{max}=30. The numbers to the left of each SAXS profile indicate the IDs of the original dataset from Grant *et al.*

Fig. 5. The computing time comparison for SHE and ZNK. The execution time of SHE is linearly proportional to the number of data points, while the computing time for ZNK method does not depend on the number of data points. When there are more than 150 data points to be computed, ZNK method has speed advantages.

Synopsis

3D Zernike polynomials are employed to efficiently compute small angle scattering profiles. 3D Zernike polynomials faithfully reproduce 3D objects with complex surfaces, including surface bound solvent. Resulting SAXS profiles are in good agreement with existing computational procedures.

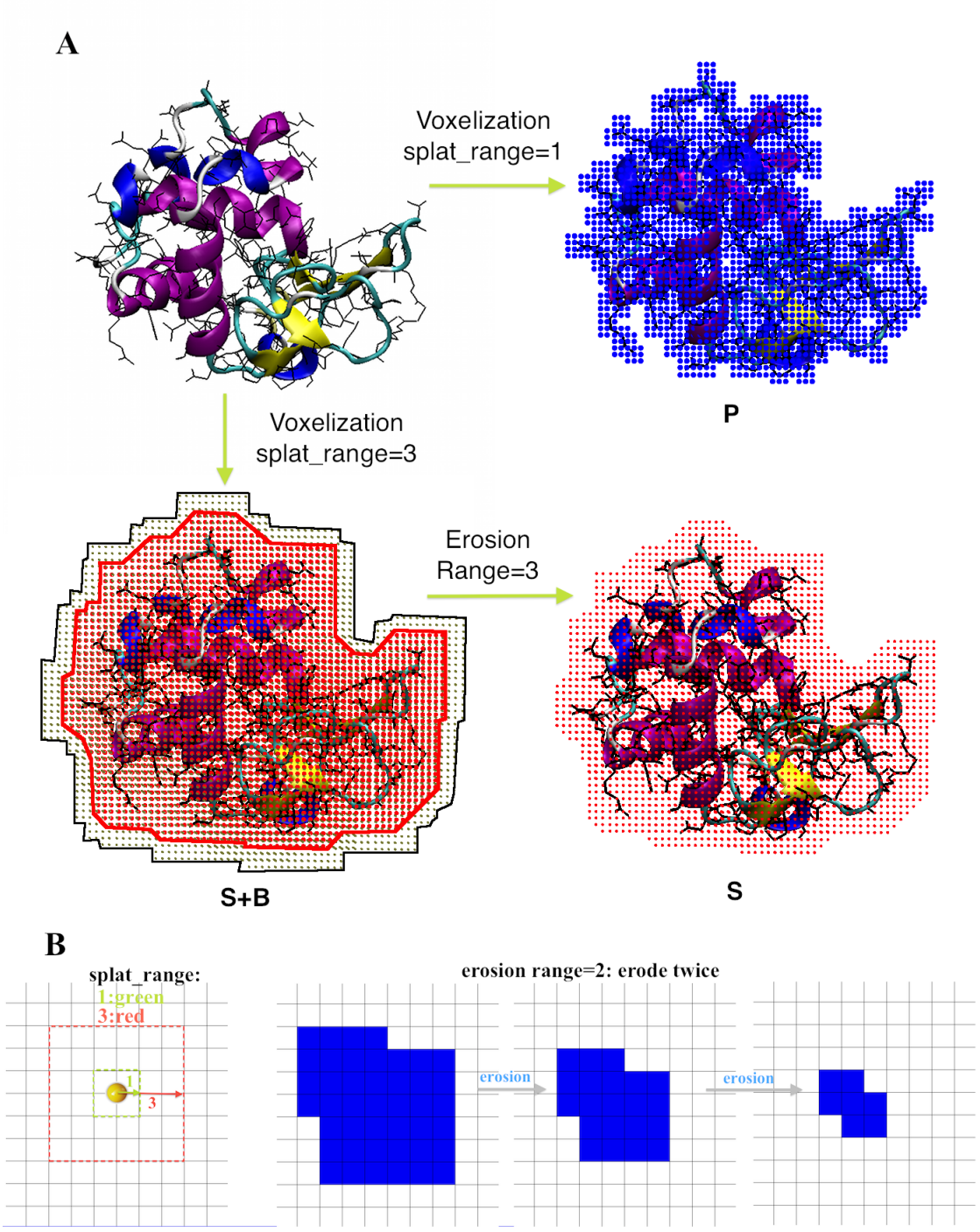

Figure 1: The voxelized representations of protein and solvents. **A**. A molecule will be mapped to 3D voxelized objects to model protein (**P**) and Solvent (**S**+**B**), and the surface bound solvent (**B**) can be decoupled from the excluded solvent (**B**) by erosion operations. **B**. The splat_range and the erosion range are defined as in the 2D scheme plot.

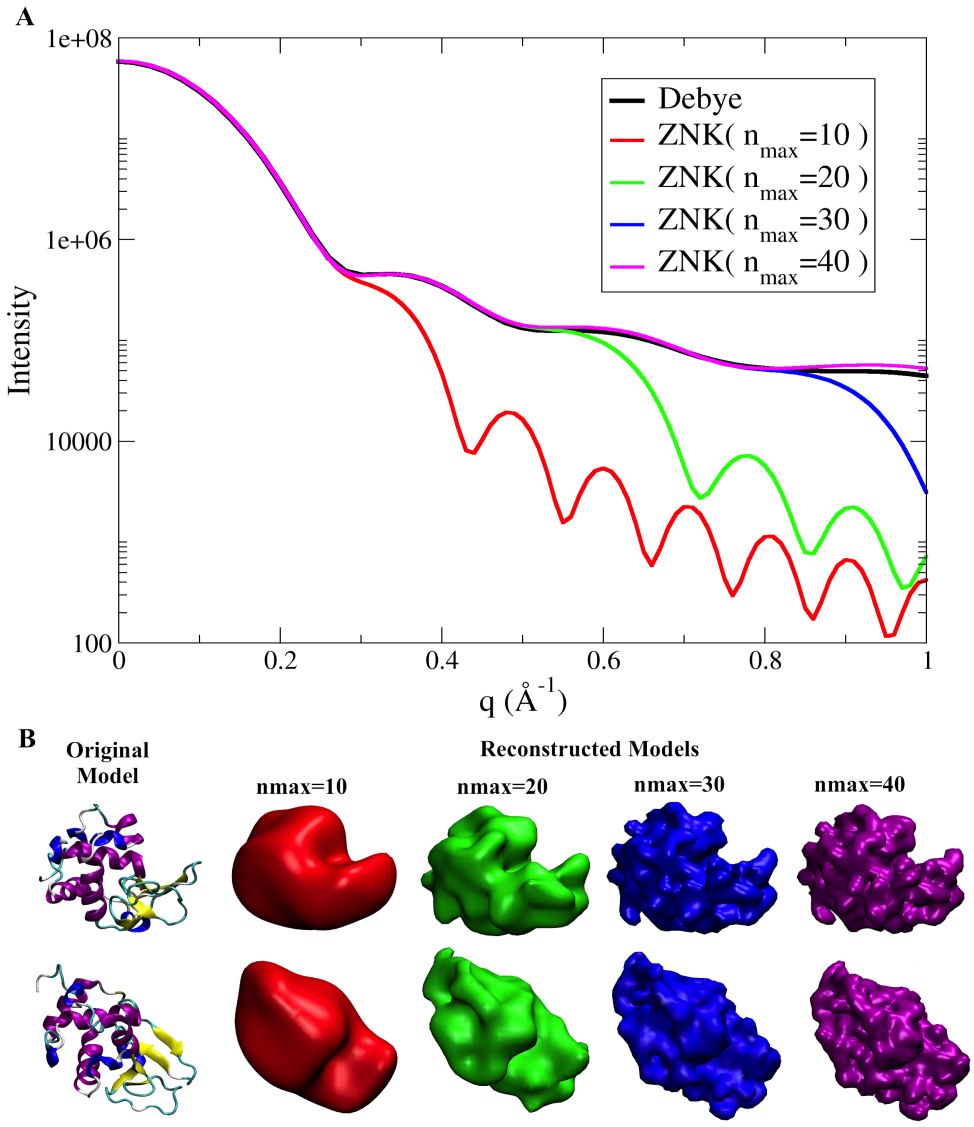

Figure 2: The SAXS profile for lysozyme *in vacua*. **A**. The resolution of the models in q-space is determined by the maximum expansion order (n_{max}): using n_{max}=40, the intensity can be accurately computed to 1.0 $\mathring{A}^{-1}$. **B**. The corresponding reconstructed models are shown at n_{max} = 10,20,30 and 40.

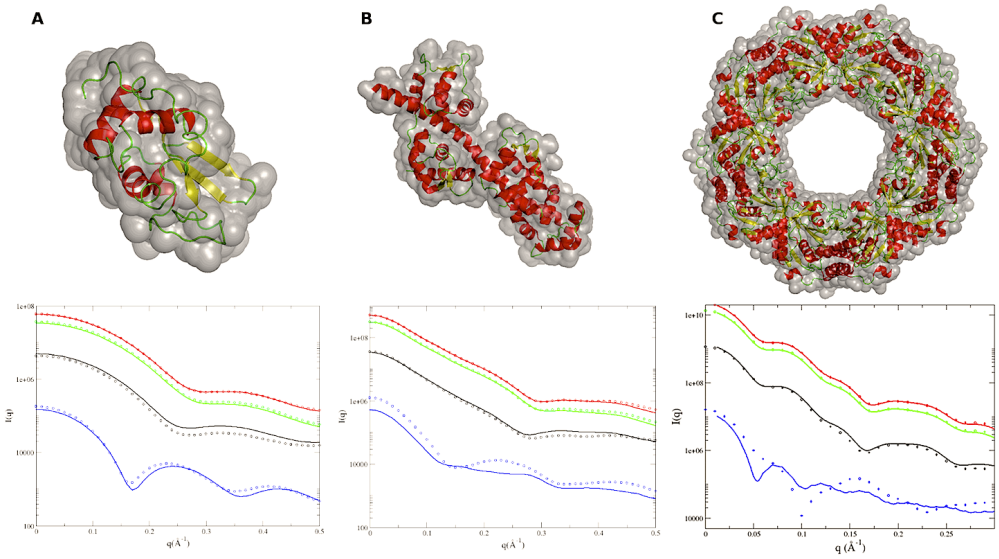

Figure 3: The SAXS profiles of individual components. The SHE (solid lines) and ZNK (open circles) method give very similar results for the overall scattering profiles (colored in black). The scattering profiles for the protein (*in vacua*) (red) and excluded solvent (green) from two methods are also in good agreements. For molecules with irregular surfaces (**B**) or large cavities/holes (**C**), the surface bound solvents have very different scattering profiles (blue). The default parameters are used: l_{max}=15 for SHE and n_{max}=30. The corresponding PDB IDs are 6LYZ(**A**), 3PN7(**B**), and 2E2G(**C**).

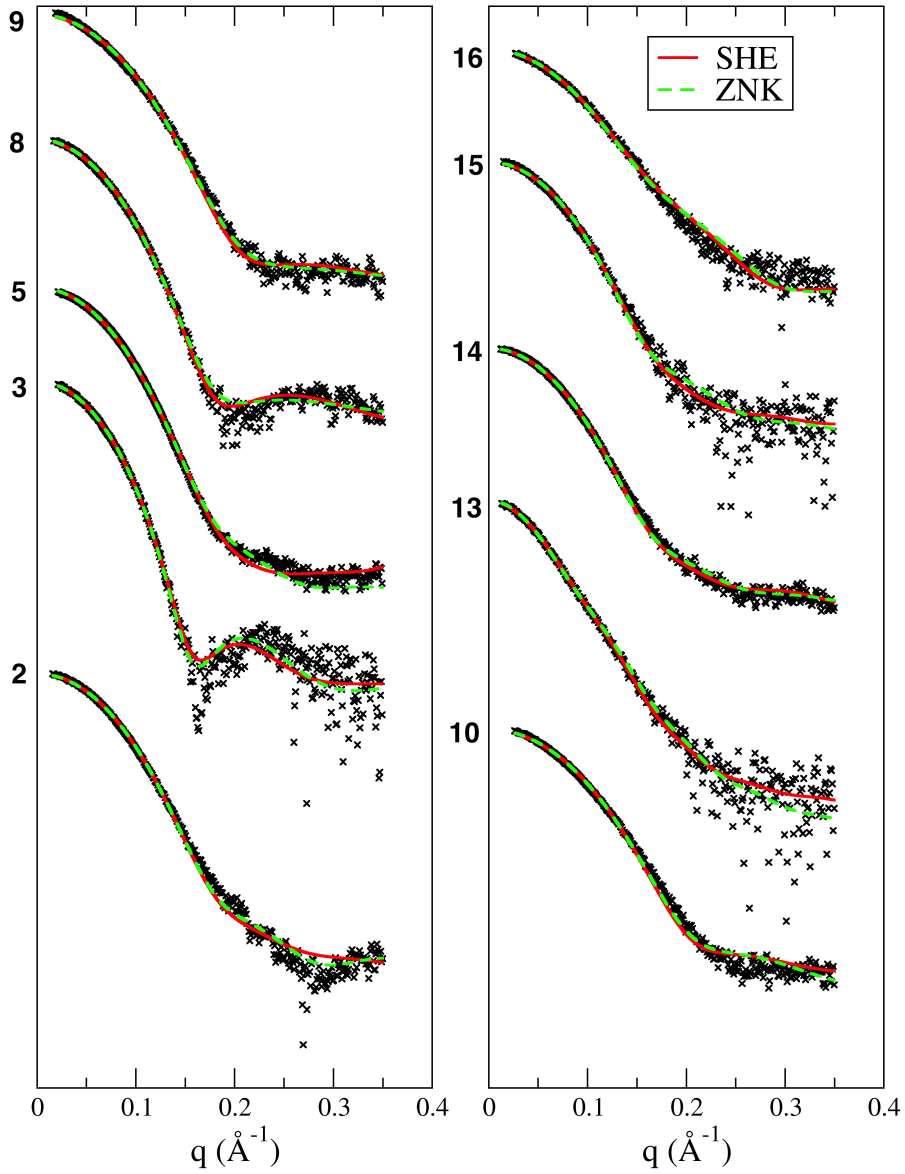

Figure 4: The fit to experimental data. The red curves are calculated using SHE method and the green curves are computed using zernike method. The default parameters are used: l_{max}=15 for SHE, and n_{max}=30. The numbers to the left of each SAXS profile indicate the IDs of the original dataset from Grant *et al.*

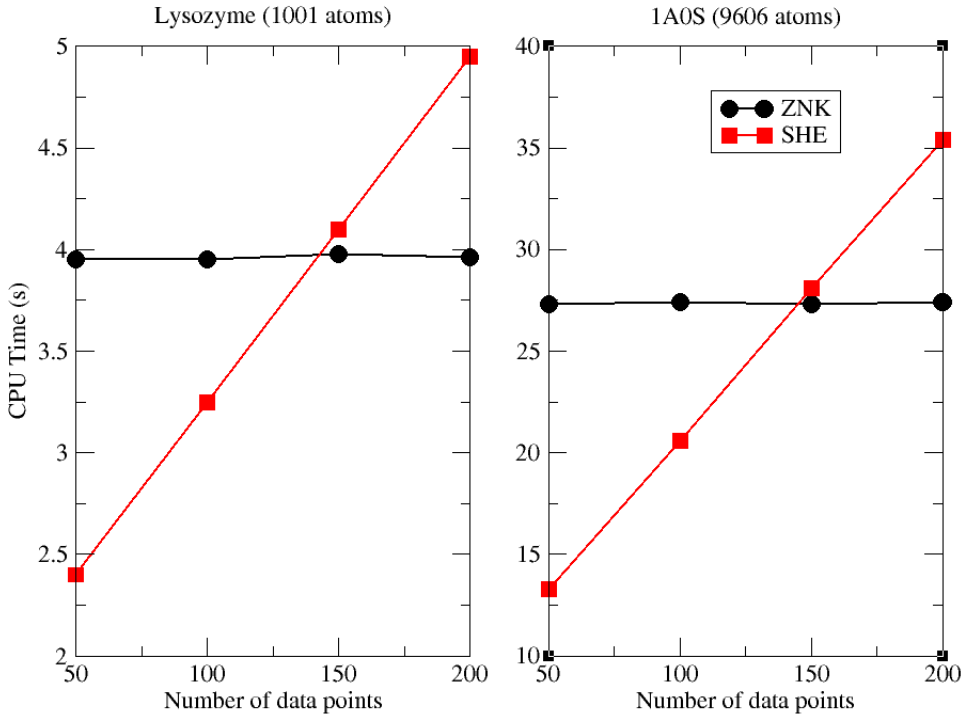

Figure 5: The computing time comparison for SHE and ZNK. The execution time of SHE is linearly proportional to the number of data points, while the computing time for ZNK method does not depend on the number of data points. When there are more than 150 data points to be computed, ZNK method has speed advantages.